A Market-Based Gittins-Type Index for a Transparent One-Armed Bandit with a Continuous Payoff Spectrum

Marcin Makowski[1], Edward Piotrowski[1]

[1]Faculty Of Physics, University Of Białystok, Białystok, Poland

We consider a sequential decision problem inspired by the multi-armed bandit problem and the Gittins index [1], in which successive observations represent incoming investment opportunities with a continuous payoff spectrum. After observing the current opportunity, the decision-maker may either accept it or reject it. Acceptance means executing the given opportunity and obtaining the corresponding payoff, whereas rejection leads to further waiting. The problem is not merely one of maximizing a single payoff, but rather of finding a decision rule that optimizes the long-run average payoff per cycle, taking into account both the quality of the accepted opportunities and the waiting time until they appear.

For the model considered here, we establish a new threshold optimality result: the problem admits an optimal solution within the class of threshold rules, and the optimal rule consists in accepting the first observation exceeding an appropriate reservation index. We then prove the existence and uniqueness of this index. Of particular importance is the fact that the optimal index has the character of a fixed point, which gives the threshold rule a transparent interpretation and facilitates its determination. The reservation index determines the minimal level of attractiveness of a market opportunity at which entering a transaction becomes optimal after taking into account both the quality of the offer and the time cost associated with capital commitment. By generalizing the problem to the case of a finite number of sources of market offers and assuming the absence of switching costs between individual sources, we show that the most important parameter describing each source of opportunities is its reservation index. As a result, the complex problem of sequential capital allocation is reduced to comparing this one-dimensional characteristic across the individual sources of market opportunities.

We also discuss a procedure for learning the reservation index under an unknown observation distribution. Since the optimal index is a fixed point of a certain average-payoff functional, its iterative estimation on the basis of successive decision cycles is both a natural and an efficient adaptive procedure. It reduces to the estimation of a single parameter by a trial-and-error method, which makes it much simpler than general reinforcement learning algorithms.

We illustrate the proposed model with the example of a bookmaker's bet under the Kelly criterion [2]. In this context, a deep connection between the reservation index and information theory emerges, since the optimal index can be interpreted as a threshold value of the Kullback–Leibler divergence between the bookmaker's distribution and the true probabilities of outcomes. This connection suggests a broader applicability of the proposed threshold rule, which may be interpreted as an information filter accepting only those opportunities for which the discrepancy between the investor's model and the market is sufficiently large.

References:

[1] J. C. Gittins, K. D. Glazebrook, and R. R. Weber, Multi-Armed Bandit Allocation Indices, 2nd ed., Wiley, 2011.

[2] E. W. Piotrowski and M. Schroeder, "Kelly criterion revisited: Optimal bets," Eur. Phys. J. B 57 (2007), 201–203.